



# Full wwPDB X-ray Structure Validation Report ⓘ

May 13, 2020 – 11:34 am BST

PDB ID : 6S41  
Title : CRYSTAL STRUCTURE OF PXR IN COMPLEX WITH XPC-7455  
Authors : Focken, T.; Maskos, K.; Griessner, A.; Krapp, S.  
Deposited on : 2019-06-26  
Resolution : 2.70 Å(reported)

This is a Full wwPDB X-ray Structure Validation Report for a publicly released PDB entry.

We welcome your comments at [validation@mail.wwpdb.org](mailto:validation@mail.wwpdb.org)

A user guide is available at

<https://www.wwpdb.org/validation/2017/XrayValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

---

The following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity : 4.02b-467  
Mogul : 1.8.5 (274361), CSD as541be (2020)  
Xtriage (Phenix) : 1.13  
EDS : 2.11  
buster-report : 1.1.7 (2018)  
Percentile statistics : 20191225.v01 (using entries in the PDB archive December 25th 2019)  
Refmac : 5.8.0158  
CCP4 : 7.0.044 (Gargrove)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : 2.11

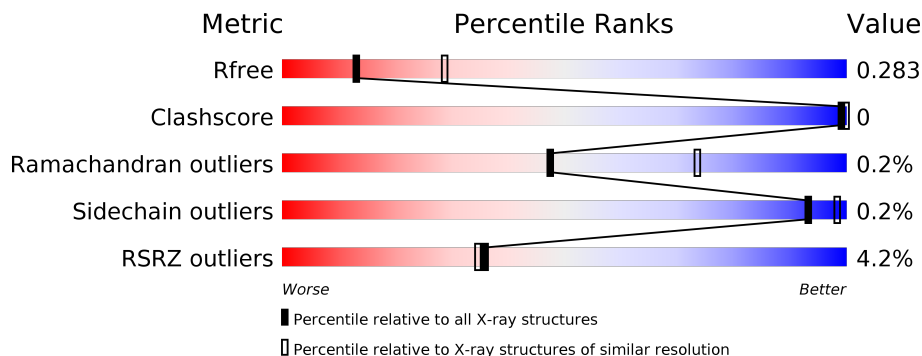
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.70 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.



Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
$R_{free}$	130704	2808 (2.70-2.70)
Clashscore	141614	3122 (2.70-2.70)
Ramachandran outliers	138981	3069 (2.70-2.70)
Sidechain outliers	138945	3069 (2.70-2.70)
RSRZ outliers	127900	2737 (2.70-2.70)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments on the lower bar indicate the fraction of residues that contain outliers for  $\geq 3$ , 2, 1 and 0 types of geometric quality criteria respectively. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions  $\leq 5\%$ . The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density. The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain
1	A	323	
1	B	323	

## 2 Entry composition i

There are 3 unique types of molecules in this entry. The entry contains 4806 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is a protein called Nuclear receptor subfamily 1 group I member 2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	284	2340	1503	402	417	18	39	1	0
1	B	293	2390	1536	410	426	18	57	0	0

There are 52 discrepancies between the modelled and reference sequences:

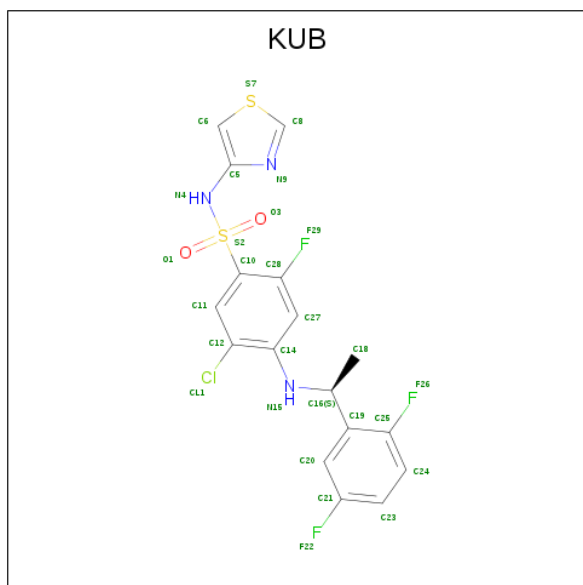
Chain	Residue	Modelled	Actual	Comment	Reference
A	673	GLY	-	expression tag	UNP O75469
A	674	GLY	-	expression tag	UNP O75469
A	675	SER	-	expression tag	UNP O75469
A	676	GLY	-	expression tag	UNP O75469
A	677	GLY	-	expression tag	UNP O75469
A	678	SER	-	expression tag	UNP O75469
A	679	SER	-	expression tag	UNP O75469
A	680	HIS	-	expression tag	UNP O75469
A	681	SER	-	expression tag	UNP O75469
A	682	SER	-	expression tag	UNP O75469
A	683	LEU	-	expression tag	UNP O75469
A	684	THR	-	expression tag	UNP O75469
A	685	GLU	-	expression tag	UNP O75469
A	686	ARG	-	expression tag	UNP O75469
A	687	HIS	-	expression tag	UNP O75469
A	688	LYS	-	expression tag	UNP O75469
A	689	ILE	-	expression tag	UNP O75469
A	690	LEU	-	expression tag	UNP O75469
A	691	HIS	-	expression tag	UNP O75469
A	692	ARG	-	expression tag	UNP O75469
A	693	LEU	-	expression tag	UNP O75469
A	694	LEU	-	expression tag	UNP O75469
A	695	GLN	-	expression tag	UNP O75469
A	696	GLU	-	expression tag	UNP O75469
A	697	GLY	-	expression tag	UNP O75469

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
A	698	SER	-	expression tag	UNP O75469
B	673	GLY	-	expression tag	UNP O75469
B	674	GLY	-	expression tag	UNP O75469
B	675	SER	-	expression tag	UNP O75469
B	676	GLY	-	expression tag	UNP O75469
B	677	GLY	-	expression tag	UNP O75469
B	678	SER	-	expression tag	UNP O75469
B	679	SER	-	expression tag	UNP O75469
B	680	HIS	-	expression tag	UNP O75469
B	681	SER	-	expression tag	UNP O75469
B	682	SER	-	expression tag	UNP O75469
B	683	LEU	-	expression tag	UNP O75469
B	684	THR	-	expression tag	UNP O75469
B	685	GLU	-	expression tag	UNP O75469
B	686	ARG	-	expression tag	UNP O75469
B	687	HIS	-	expression tag	UNP O75469
B	688	LYS	-	expression tag	UNP O75469
B	689	ILE	-	expression tag	UNP O75469
B	690	LEU	-	expression tag	UNP O75469
B	691	HIS	-	expression tag	UNP O75469
B	692	ARG	-	expression tag	UNP O75469
B	693	LEU	-	expression tag	UNP O75469
B	694	LEU	-	expression tag	UNP O75469
B	695	GLN	-	expression tag	UNP O75469
B	696	GLU	-	expression tag	UNP O75469
B	697	GLY	-	expression tag	UNP O75469
B	698	SER	-	expression tag	UNP O75469

- Molecule 2 is 4-[[[(1 {S})-1-[2,5-bis(fluoranyl)phenyl]ethyl]amino]-5-chloranyl-2-fluoranyl- {N}-(1,3-thiazol-4-yl)benzenesulfonamide (three-letter code: KUB) (formula: C<sub>17</sub>H<sub>13</sub>ClF<sub>3</sub>N<sub>3</sub>O<sub>2</sub>S<sub>2</sub>) (labeled as "Ligand of Interest" by author).



Mol	Chain	Residues	Atoms						ZeroOcc	AltConf		
			Total	C	Cl	F	N	O			S	
2	A	1	Total	28	17	1	3	3	2	2	0	0
2	B	1	Total	28	17	1	3	3	2	2	0	0

- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	10	Total	O	0	0
			10	10		
3	B	10	Total	O	0	0
			10	10		



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	84.32Å 89.00Å 105.81Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	68.11 – 2.70 45.48 – 2.70	Depositor EDS
% Data completeness (in resolution range)	96.9 (68.11-2.70) 97.0 (45.48-2.70)	Depositor EDS
$R_{merge}$	0.09	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.13 (at 2.69Å)	Xtrriage
Refinement program	REFMAC 5.8.0155	Depositor
R, $R_{free}$	0.213 , 0.259 0.235 , 0.283	Depositor DCC
$R_{free}$ test set	784 reflections (3.60%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	43.4	Xtrriage
Anisotropy	0.156	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.31 , 23.6	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.91	EDS
Total number of atoms	4806	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	32.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.61% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical values of  $\langle |L| \rangle$ ,  $\langle L^2 \rangle$  for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

## 5 Model quality [i](#)

### 5.1 Standard geometry [i](#)

Bond lengths and bond angles in the following residue types are not validated in this section: KUB

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with  $|Z| > 5$  is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	# Z  >5	RMSZ	# Z  >5
1	A	0.59	2/2390 (0.1%)	0.79	4/3216 (0.1%)
1	B	0.68	4/2440 (0.2%)	0.82	8/3281 (0.2%)
All	All	0.64	6/4830 (0.1%)	0.81	12/6497 (0.2%)

All (6) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	140	VAL	CB-CG2	-11.34	1.29	1.52
1	B	140	VAL	CB-CG1	9.42	1.72	1.52
1	B	319	LEU	CG-CD2	-5.99	1.29	1.51
1	B	319	LEU	CG-CD1	5.51	1.72	1.51
1	A	311	THR	CB-OG1	-5.30	1.32	1.43
1	A	317	GLN	CG-CD	-5.01	1.39	1.51

All (12) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	140	VAL	CA-CB-CG2	10.49	126.64	110.90
1	B	319	LEU	CB-CG-CD2	8.10	124.76	111.00
1	B	140	VAL	CA-CB-CG1	-6.22	101.58	110.90
1	B	319	LEU	CB-CG-CD1	-6.08	100.66	111.00
1	B	413	ARG	NE-CZ-NH1	5.81	123.20	120.30
1	A	413	ARG	NE-CZ-NH1	5.71	123.16	120.30
1	A	319	LEU	CB-CG-CD2	5.58	120.48	111.00
1	A	203	ARG	NE-CZ-NH1	5.45	123.03	120.30
1	B	203	ARG	NE-CZ-NH1	5.36	122.98	120.30
1	B	686	ARG	NE-CZ-NH1	5.28	122.94	120.30
1	A	360	ARG	NE-CZ-NH1	5.19	122.90	120.30
1	B	162	PHE	CB-CG-CD1	5.07	124.35	120.80



There are no chirality outliers.

There are no planarity outliers.

## 5.2 Too-close contacts

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	2340	0	2354	1	0
1	B	2390	0	2407	2	0
2	A	28	0	0	0	0
2	B	28	0	0	0	0
3	A	10	0	0	0	0
3	B	10	0	0	0	0
All	All	4806	0	4761	3	0

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). The all-atom clashscore for this structure is 0.

All (3) close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:254:ILE:HD12	1:B:283:LEU:HB3	1.91	0.52
1:A:254:ILE:HD12	1:A:283:LEU:HB3	2.01	0.42
1:B:279:ALA:HB1	1:B:283:LEU:HD13	2.03	0.40

There are no symmetry-related clashes.

## 5.3 Torsion angles

### 5.3.1 Protein backbone

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the backbone conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	A	277/323 (86%)	271 (98%)	5 (2%)	1 (0%)	34	60
1	B	285/323 (88%)	282 (99%)	3 (1%)	0	100	100
All	All	562/646 (87%)	553 (98%)	8 (1%)	1 (0%)	47	73

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	319	LEU

### 5.3.2 Protein sidechains [i](#)

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the sidechain conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	A	259/284 (91%)	258 (100%)	1 (0%)	91	97
1	B	263/284 (93%)	263 (100%)	0	100	100
All	All	522/568 (92%)	521 (100%)	1 (0%)	93	98

All (1) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	238	SER

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

### 5.3.3 RNA [i](#)

There are no RNA molecules in this entry.

## 5.4 Non-standard residues in protein, DNA, RNA chains [i](#)

There are no non-standard protein/DNA/RNA residues in this entry.

## 5.5 Carbohydrates [i](#)

There are no carbohydrates in this entry.

## 5.6 Ligand geometry [i](#)

2 ligands are modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with  $|Z| > 2$  is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z  > 2	Counts	RMSZ	# Z  > 2
2	KUB	B	701	-	27,30,30	1.26	2 (7%)	34,44,44	2.16	10 (29%)
2	KUB	A	701	-	27,30,30	1.29	2 (7%)	34,44,44	2.04	8 (23%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	KUB	B	701	-	-	5/17/19/19	0/3/3/3
2	KUB	A	701	-	-	1/17/19/19	0/3/3/3

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	701	KUB	C6-S7	3.65	1.75	1.70
2	A	701	KUB	C6-S7	3.36	1.75	1.70
2	B	701	KUB	O1-S2	-3.04	1.40	1.43
2	A	701	KUB	C28-C10	2.43	1.40	1.39

All (18) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	A	701	KUB	C10-S2-N4	7.34	115.84	107.27
2	B	701	KUB	C10-S2-N4	5.86	114.11	107.27

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	A	701	KUB	C12-C14-N15	-5.28	117.92	120.43
2	B	701	KUB	C12-C14-N15	-4.69	118.20	120.43
2	B	701	KUB	O1-S2-N4	-4.00	96.72	106.73
2	B	701	KUB	C27-C28-C10	-3.91	120.83	123.10
2	B	701	KUB	C20-C19-C25	3.57	119.45	116.48
2	B	701	KUB	C24-C25-C19	-3.27	120.08	123.72
2	A	701	KUB	C20-C19-C25	3.09	119.05	116.48
2	A	701	KUB	C24-C25-C19	-2.77	120.64	123.72
2	B	701	KUB	C23-C21-C20	-2.51	120.03	123.29
2	A	701	KUB	O1-S2-N4	-2.43	100.66	106.73
2	B	701	KUB	C11-C10-S2	2.35	122.64	118.51
2	A	701	KUB	C27-C28-C10	-2.34	121.74	123.10
2	A	701	KUB	C23-C21-C20	-2.31	120.30	123.29
2	B	701	KUB	C11-C12-C14	-2.21	119.91	121.72
2	B	701	KUB	F26-C25-C19	2.04	121.53	118.23
2	A	701	KUB	C11-C10-S2	2.01	122.04	118.51

There are no chirality outliers.

All (6) torsion outliers are listed below:

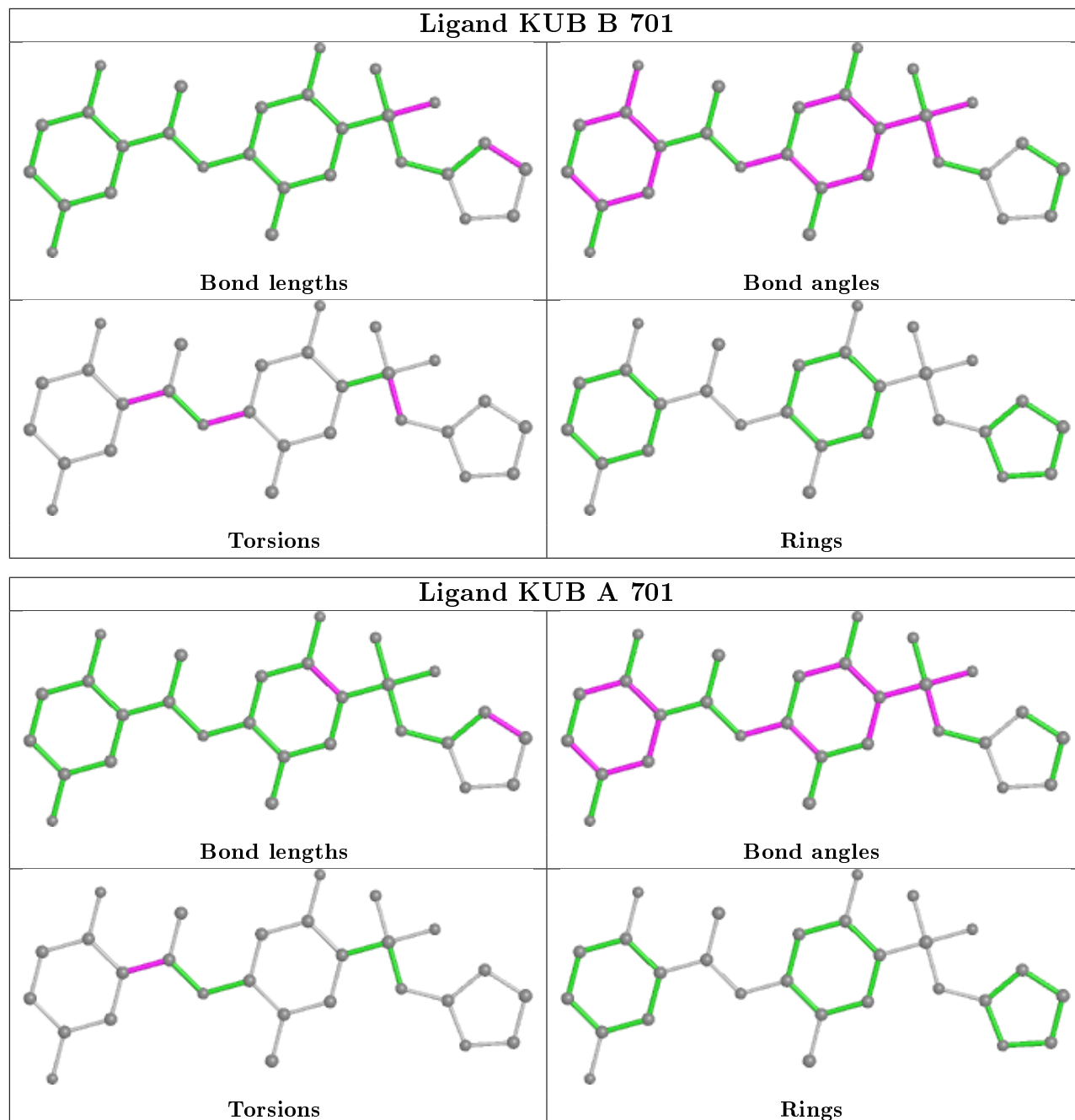
Mol	Chain	Res	Type	Atoms
2	B	701	KUB	C5-N4-S2-O1
2	B	701	KUB	C5-N4-S2-C10
2	B	701	KUB	C5-N4-S2-O3
2	A	701	KUB	N15-C16-C19-C20
2	B	701	KUB	C12-C14-N15-C16
2	B	701	KUB	N15-C16-C19-C20

There are no ring outliers.

No monomer is involved in short contacts.

The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the validation Tables will also be included. For torsion angles, if less than 5% of the Mogul distribution of torsion angles is within 10 degrees of the torsion angle in question, then that torsion angle is considered an outlier. Any bond that is central to one or more torsion angles identified as an outlier by Mogul will be highlighted in the graph. For rings, the root-mean-square deviation (RMSD) between the ring in question and similar rings identified by Mogul is calculated over all ring torsion angles. If the average RMSD is greater than 60 degrees and the minimal RMSD between the ring in question and any Mogul-identified rings is also greater than 60 degrees, then that ring is considered an outlier. The outliers are highlighted in purple. The color gray indicates Mogul did not find sufficient

equivalents in the CSD to analyse the geometry.



## 5.7 Other polymers [i](#)

There are no such residues in this entry.

## 5.8 Polymer linkage issues [i](#)

There are no chain breaks in this entry.

## 6 Fit of model and data

### 6.1 Protein, DNA and RNA chains

In the following table, the column labelled ‘#RSRZ> 2’ contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95<sup>th</sup> percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled ‘Q< 0.9’ lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ>2	OWAB(Å <sup>2</sup> )	Q<0.9
1	A	284/323 (87%)	0.03	6 (2%) 63 65	14, 28, 52, 83	15 (5%)
1	B	293/323 (90%)	0.23	18 (6%) 21 20	13, 27, 71, 95	17 (5%)
All	All	577/646 (89%)	0.13	24 (4%) 36 35	13, 27, 63, 95	32 (5%)

All (24) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	196	ALA	5.8
1	B	177	VAL	5.3
1	B	698	SER	3.9
1	B	314	GLY	3.5
1	B	197	ALA	3.3
1	B	218	GLU	3.2
1	B	683	LEU	3.1
1	B	384	PRO	3.1
1	B	697	GLY	2.7
1	B	207	CYS	2.6
1	A	231	SER	2.6
1	B	199	TRP	2.5
1	B	417	ILE	2.4
1	A	318	LEU	2.4
1	A	194	GLU	2.3
1	B	232	GLY	2.3
1	B	178	LEU	2.2
1	B	231	SER	2.2
1	A	232	GLY	2.2
1	B	315	PHE	2.2
1	B	201	GLN	2.1
1	A	208	SER	2.1
1	B	416	ASP	2.1
1	A	311	THR	2.0

## 6.2 Non-standard residues in protein, DNA, RNA chains [i](#)

There are no non-standard protein/DNA/RNA residues in this entry.

## 6.3 Carbohydrates [i](#)

There are no carbohydrates in this entry.

## 6.4 Ligands [i](#)

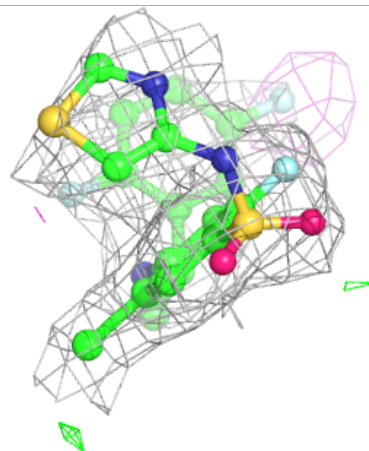
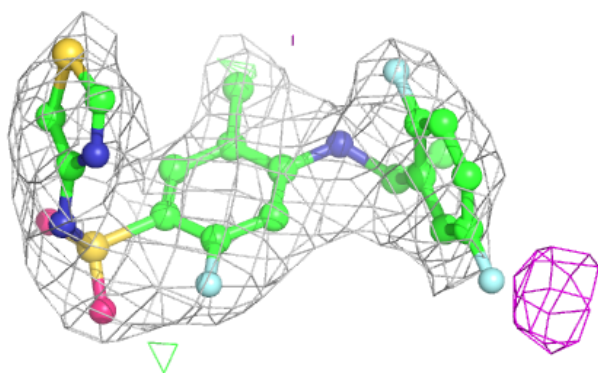
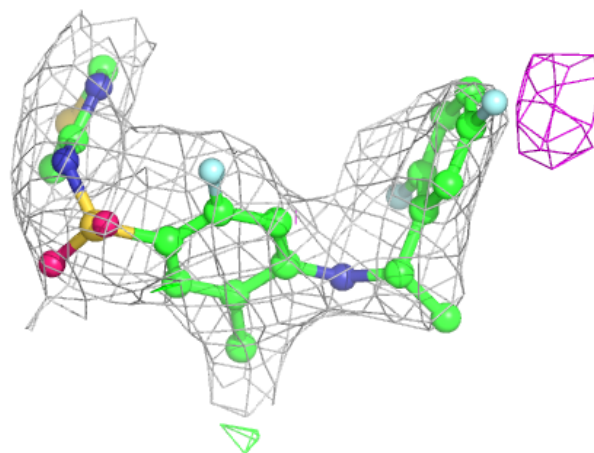
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95<sup>th</sup> percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors( $\text{\AA}^2$ )	Q<0.9
2	KUB	B	701	28/28	0.91	0.23	57,74,90,111	0
2	KUB	A	701	28/28	0.93	0.19	48,56,65,75	0

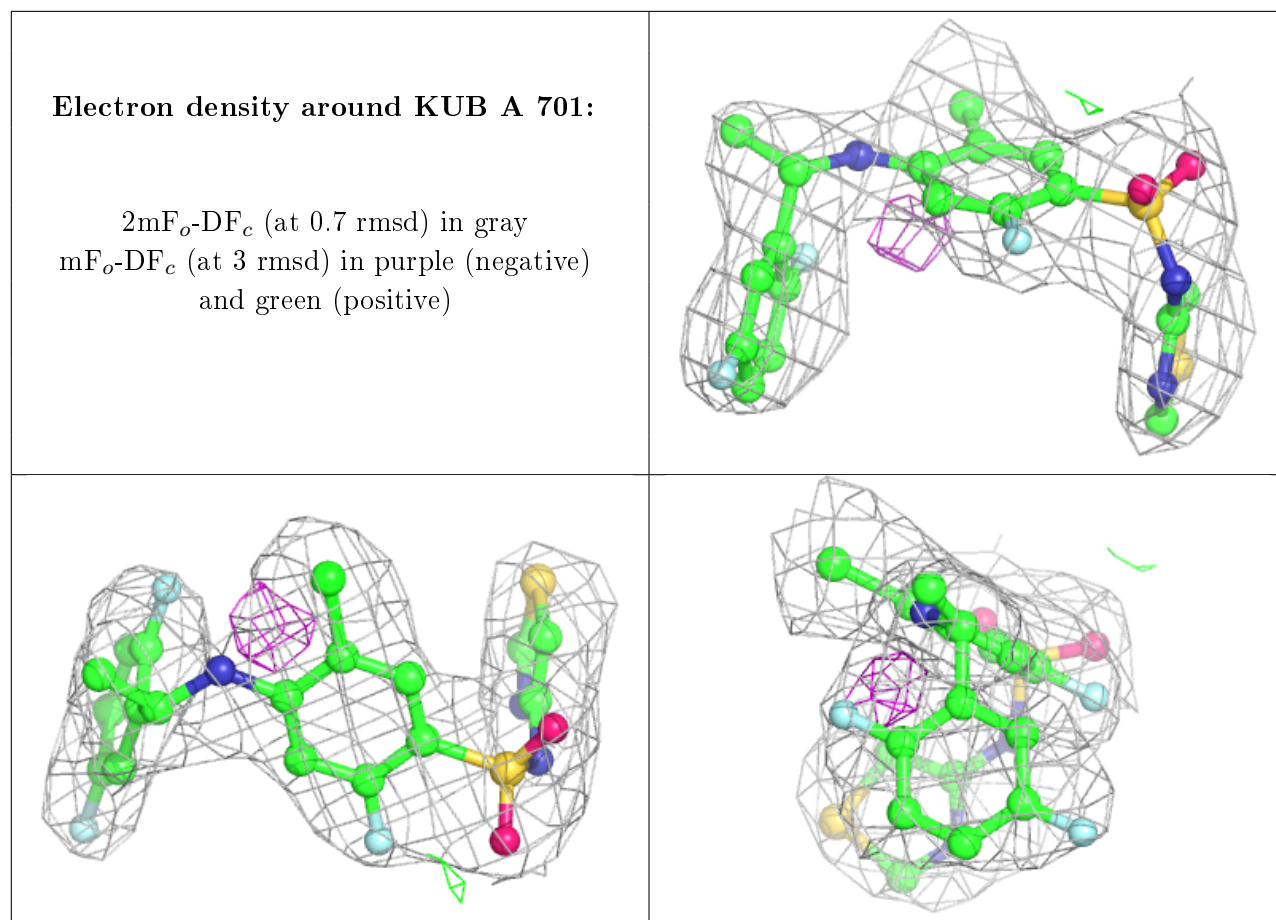
The following is a graphical depiction of the model fit to experimental electron density of all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the geometry validation Tables will also be included. Each fit is shown from different orientation to approximate a three-dimensional view.

**Electron density around KUB B 701:**

$2mF_o-DF_c$  (at 0.7 rmsd) in gray  
 $mF_o-DF_c$  (at 3 rmsd) in purple (negative)  
and green (positive)







## 6.5 Other polymers [i](#)

There are no such residues in this entry.